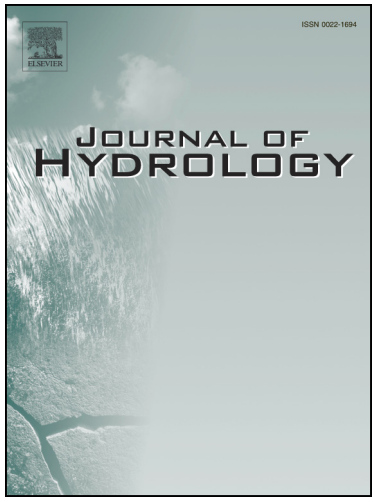
Quantification of prediction uncertainty using imperfect subsurface models with model error estimation



Quantification of prediction uncertainty using imperfect subsurface models with model error estimation

Muzammil Hussain Rammay, Ahmed H. Elsheikh, Yan Chen

PII: S0022-1694(19)30213-6
DOI: https://doi.org/10.1016/j.jhydrol.2019.02.056
Reference: HYDROL 23530

To appear in: *Journal of Hydrology*

Please cite this article as: Rammay, M.H., Elsheikh, A.H., Chen, Y., Quantification of prediction uncertainty using imperfect subsurface models with model error estimation, *Journal of Hydrology* (2019), doi: https://doi.org/10.1016/j.jhydrol.2019.02.056

Quantification of prediction uncertainty using imperfect subsurface models with model error estimation

Muzammil Hussain Rammay[1], Ahmed H. Elsheikh[1], and Yan Chen[2]

[1]Heriot–Watt University, UK

[2]Geoscience Research Centre, Total E&P UK

March 3, 2019

Abstract

Subsurface reservoirs are far more heterogeneous and complex than the simulation models in terms of scale, assumptions and description. In this work, we address the issue of prediction reliability while calibrating imperfect/low-fidelity reservoir models. The main goal is to avoid over-confident and inaccurate predictions by including a model for the bias terms (i.e. error-model of a predefined form) during the history matching process. Our aim is to obtain unbiased posterior distributions of the physical model parameters and thus improving the prediction capacity of the calibrated low-fidelity reservoir models. We formulate the parameter estimation problem as a joint estimation of the imperfect model parameters and the error-model parameters. The structure of the error-model and the prior distributions of the error-model parameters are evaluated before calibration through analysis of leading sources of the modeling errors. We adopt a Bayesian framework for solving the inverse problem, where we utilize the ensemble smoother with multiple data assimilation (ES-MDA) as a practical history matching algorithm.

We provide two test cases, where the impact of typical model errors originating from grid coarsening/up-scaling and from utilizing an imperfect geological model description is investigated. For both cases results from the ES-MDA update with and without accounting for model error are compared in terms of estimated physical model parameters, quality of match to historical data and forecasting ability compared to held out data. The test results show that calibration

of the imperfect physical model without accounting for model errors results in extreme values of the calibrated model parameters and a biased posterior distribution. With accounting for modeling errors the posterior distribution of the model parameters is less biased (i.e. nearly unbiased) and improved forecasting skills with higher prediction accuracy/reliability is observed. Moreover, the consistency between the different runs of the ES-MDA is improved by including the modeling error component. Although the examples in the paper consider the oil-water system with permeabilities being parameters of the physical model, the developed methodology is general and can be applied to typical ground water hydrology models.

1 Introduction

In subsurface reservoir modeling, various approximations are introduced at different stages of the modeling process which in turn render most of the models to be imperfect and low-fidelity in nature. However, these imperfect models are generally still useful for understanding the key physical interactions within the subsurface regions of interest. The sources of approximations (a.k.a. modeling errors) include: properties up-scaling (grid coarsening), discretization errors, imperfect reservoir fluid properties, relative permeability, reservoir geology description/parameterization and approximate representation of the complete complex subsurface fluid flow physics (e.g. black-oil model in place of a compositional model or constant rock compressibility assumption).

In the context of error modeling, grid up-scaling has been widely studied within the reservoir simulation community. In the published literature, there exist a number of efficient up-scaling techniques (Durlofsky, 2003), aiming to obtain optimal upscaled properties. However, up-scaling errors are not completely eliminated by most of these methods. Discretization errors also cannot be eliminated (Ertekin et al., 2001), even after selecting an optimal grid size and utilizing adaptive time stepping techniques. Additionally, numerical simulation using an optimal fine-grid could be computationally prohibitive especially for tasks that typically requires many simulation runs, for example history matching or robust optimization problems. Various techniques have been proposed to address this computational bottleneck, for example reduced order modeling and proxy models among many other techniques [c.f., Silva et al., 2007; Rammay and Abdulraheem, 2014; Cardoso

et al., 2009]. An alternative approach is to utilize an upscaled model instead of the fine scale model for multi-query computationally demanding tasks (e.g. uncertainty quantification problems). In the context of history matching, if the up-scaling errors are not modeled during the parameter inference step, the posterior distributions of the model parameters is likely to be biased and this bias will subsequently affect the future predictions of the engineering quantities of interest (e.g. oil, gas and water rates/pressure). Omre et al. (2004) approximated the up-scaling and discretization errors by computing samples or realizations of the error data using pairs of fine- and coarse-scale models. The model errors due to up-scaling were then estimated using a multiple regression technique and added to the coarse scale model predictions during the history matching process. Lodoen et al. (2005) utilized a similar procedure on a different set of test cases while employing a more accurate up-scaling procedure.

Accurate reservoir geology description is another challenging task due to various uncertainties including: channel geometry, faults shape, facies proportion, stratigraphic and/or structural frameworks. It is widely known that unrealistic geological models could be calibrated to match the historical data [c.f., Carter et al., 2006; Refsgaard et al., 2012]. However, these fitted models fail to provide reliable predictions and could ultimately mislead the reservoir development plans [c.f., Carter et al., 2006; Refsgaard et al., 2012]. Although considerable effort is often put into constructing geological models that are as realistic as possible, it is very difficult to maintain this geological realism while updating them to match the observed data (Sun and Durlofsky, 2017).

Accounting for model errors during the calibration process has attracted a large body of research [c.f., Oliver and Alfonzo, 2018; Dreano et al., 2017; Josset et al., 2015], where various approaches have been developed to account for the model-error component during model calibration [c.f., Hansen et al., 2014; Evin et al., 2014; Reichert and Schuwirth, 2012]. These approaches vary according to the different behavior and complexity of the modeled physical system. For example in hydro-geophysical systems, Köpke et al. (2017) accounted for the model-error component using orthonormal basis generated from an error dictionary which is continuously enriched during the calibration process. The models of the bias or error component could be generally classified as either input dependent (Giudice et al., 2013) or output dependent (Evin et al., 2014). Input

dependent model error formulation represents the error components as a function of the model parameters. For example, a reservoir model-error can vary with permeability realizations or other input model parameters. O'Sullivan and Christie (2005) utilized an input dependent formulation for the model-error where the authors computed model-error realizations using the difference between a fine-grid and coarse-grid model outputs. During the calibration process, an interpolation of the error component was performed to estimate a correction term to the coarse-grid model predictions. Lødøen and Tjelmeland (2010) used multiple linear regression algorithm to model errors, where the residual part for the multiple regression was assumed to depend on the model input parameters. The residual terms were modeled using a zero mean Gaussian Process. Giudice et al. (2013) used an input dependent model-error representation to improve uncertainty estimation in urban hydrological models. In that application, the model error variance was set to be depending on the input of the rainfall term. Output dependent model error formulation represents the error components as a function of the output of the physical model. For example, Evin et al. (2014) utilized an output dependent formulation for the model-error heteroscedasticity as a function of the simulated streamflow. In a realistic setting, where large models are utilized (i.e. millions of input parameters), it is hard to relate the model errors to the high dimensional input parameter space and output dependent or input/output (I/O) independent forms of the model-error were proposed as an alternative approach that might have some advantages over the input-dependent error-models (Giudice et al., 2013).

Model-error representation can also be classified as either an external bias description (EBD) or an internal noise description (IND) (Giudice et al., 2015). EBD was developed using the background of statistical inference in a regression type framework (Giudice et al., 2015). In EBD, the model-error term is added externally into the forward model (approximate or inadequate model) output. In IND, the model-error is formulated as an additional term of the state space (Giudice et al., 2015). This approach is also known as state space modelling or stochastic gray-box modelling [c.f., Moradkhani et al., 2012; Kristensen et al., 2004]. Giudice et al. (2015) concluded that EBD has some advantages over IND in terms of long-term predictions.

In this paper, we utilize an EBD and I/O independent error-model formulation as it is more

suitable for large scale models (e.g. subsurface oil and gas reservoir models) relying on black-box simulators. Several typical sources of model errors are present in the test cases investigated in this paper, including a coarse grid, less detailed geological representation (i.e. upscaling of different types of geological features including variogram based and channelized geology), discretization errors and slight change in well locations due to grid coarsening. The error-model formulation presented here, assumes that the total modeling errors consists of two components: structural component and noise-like component. We note that the structural component is often neglected in the bias correction approaches developed in hydrological literature [c.f., Maier et al., 2014; Vrugt, 2016; White et al., 2014]. In this study prior to model calibration, the structure of the model error is estimated and represented using several basis functions, and the magnitude of the noise-like component is quantified. During history matching, the weights of the basis functions are jointly calibrated with the physical model parameters using data observed at well locations. The noise-like part of the model error is also accounted for during the history matching process to avoid over-fitting of the error model. We note that the presented formulation is general and can be applied to other sources of modeling errors when dealing with low-fidelity physical models. The low-fidelity models are generally used as efficient surrogate models for computationally demanding tasks [c.f., Asher et al., 2015; Laloy et al., 2013].

For the Bayesian inversion, we use a particular type of iterative ensemble smoother ES-MDA (Emerick and Reynolds, 2013). The formulation of ES-MDA has some similarities with Kalman filtering algorithms (Sun et al., 2016). However, ES-MDA assimilates data from different times simultaneously and the same set of data is assimilated multiple times with an inflated data noise covariance matrix which is equivalent to annealing approaches (Stordal and Elsheikh, 2015). The rest of the paper is organized as follows: In Section 2, we present some background on Bayesian inverse modeling followed by the proposed error-model formulation. Following that, we present the case studies in Section 3. The results of the case studies are discussed in Section 4 followed by the conclusions of our work in Section 5.

2 Methodology

Bayesian inverse modeling is a generic inference framework that is widely adopted for calibration of reservoir models while accounting for different types/sources of uncertainties. In the Bayesian framework, the conditional probabilities $p(\mathbf{m}|\mathbf{d}_{obs})$ of the model parameters $\mathbf{m}$ given the observational data $\mathbf{d}_{obs}$ (a.k.a. posterior distribution of the model parameters) is estimated using Bayes rule (Oliver et al., 2008):

$$p(\mathbf{m}|\mathbf{d}_{obs}) \propto p(\mathbf{d}_{obs}|\mathbf{m})\ p(\mathbf{m}), \tag{1}$$

where $\mathbf{m}$ is the model parameters vector of size N_m, $\mathbf{d}_{obs}$ is the observations vector of size N_d, $p(\mathbf{m})$ is the prior probability of the model parameters and $p(\mathbf{d}_{obs}|\mathbf{m})$ is the likelihood function of the data given a specific realization of the model parameters $\mathbf{m}$. It is common to assume a Gaussian prior:

$$p(\mathbf{m}) \propto \exp\big(-\frac{1}{2}(\mathbf{m}-\mathbf{m}_{pr})^{\intercal}\mathbf{C}_M^{-1}(\mathbf{m}-\mathbf{m}_{pr})\big), \tag{2}$$

where $\mathbf{m}_{pr}$ is an N_m dimensional vector of the mean prior model parameters and $\mathbf{C}_M$ is the covariance matrix of the prior model parameters. It is also common to assume that data noise is Gaussian, so that the likelihood function takes the form:

$$p(\mathbf{d}_{obs}|\mathbf{m}) \propto \exp\big(-\frac{1}{2}(\mathbf{d}_{obs}-\mathbf{d})^{\intercal}\mathbf{C}_D^{-1}(\mathbf{d}_{obs}-\mathbf{d})\big), \tag{3}$$

where $\mathbf{d}$ is the simulated or predicted data vector using the model parameters $\mathbf{m}$ and $\mathbf{C}_D$ is the error/noise covariance matrix which is defined in Sect. 2.1 and 2.2 depending on the utilized history matching procedure. Using these definitions, Bayes' rule defined in Eq. 1 could be expanded as following:

$$p(\mathbf{m}|\mathbf{d}_{obs}) \propto \exp\bigg(-\frac{1}{2}\big((\mathbf{d}_{obs}-\mathbf{d})^{\intercal}\mathbf{C}_D^{-1}(\mathbf{d}_{obs}-\mathbf{d}) + (\mathbf{m}-\mathbf{m}_{pr})^{\intercal}\mathbf{C}_M^{-1}(\mathbf{m}-\mathbf{m}_{pr})\big)\bigg). \tag{4}$$

Several algorithms could be used to generate samples from the posterior distribution of the model parameters (Oliver et al., 2008). Among those Markov Chain Monte Carlo (MCMC) is

an exact method for sampling. However, MCMC can be computationally expensive due to the large number of iterations needed to reach convergence and the sequential nature of the method. Ensemble-based methods have been widely used for calibrating subsurface flow models due to the computational feasibility and parallel nature of ensemble methods. In this study, we utilize the ensemble smoother with multiple data assimilation (ES-MDA) algorithm for the calibration step (Emerick and Reynolds, 2013). ES-MDA belongs to a class of iterative ensemble smoothing techniques that could be used to solve non-linear inverse problem iteratively with an inflated noise covariance matrix. The ES-MDA algorithm steps are summarized as follows:

- Select the number of iterations (number of data assimilation) N_a and the inflation factor α. A common choice of the inflation factor is to set it as a constant value for all iterations $\alpha = N_a$
- Initialize an ensemble of model parameters and perturb the observation data for each ensemble member using:

$$\mathbf{d}_{uc,j} = \mathbf{d}_{obs} + \sqrt{\alpha}\, \mathbf{C}_D^{1/2}\, \mathbf{z}_d, \tag{5}$$

  where the subscript j is the ensemble member index $j = 1 \dots N_e$ and N_e is the ensemble size, $\mathbf{d}_{uc,j}$ is N_d dimensional vector of perturbed observation, $\mathbf{z}_d$ is the N_d dimensional vector with standard Gaussian random variables as its components (i.e. $\mathbf{z}_d \sim \mathcal{N}(0, \mathbf{I}_{N_d,N_d})$).
- Update each ensemble member using,

$$\mathbf{m}_j^{(i+1)} = \mathbf{m}_j^{(i)} + \mathbf{C}_{MD}\, (\mathbf{C}_{DD} + \alpha\, \mathbf{C}_D)^{-1} (\mathbf{d}_{uc,j} - \mathbf{d}_j), \tag{6}$$

  where the superscript i is the iteration index, $\mathbf{C}_{DD}$ is the model output covariance matrix and $\mathbf{C}_{MD}$ is the cross covariance matrix of model parameters and model predictions.
- Repeat the above steps for all iterations, from $i = 1$ to N_a

2.1 Procedures for history matching of reservoir models

In this study, two types of history matching procedures are investigated: history matching while neglecting model-discrepancy (i.e. standard history matching procedure) and joint history matching

of the model parameters and the parameters of an error/bias model. In this paper, we use the term model-discrepancy and model-error interchangeably. The standard history matching procedure relies on an implicit assumption that the model-errors are generally small and could be neglected (i.e. the simulation model is perfect). Mathematically, if an accurate/high-fidelity model is utilized, the observed data is formulated as (Giudice et al., 2013):

$$\mathbf{d}_{obs} = g(\mathbf{m}_{true}) + \boldsymbol{\epsilon}_d, \tag{7}$$

where $g(.)$ is a nonlinear function representing the accurate/high-fidelity forward simulation model, $\mathbf{m}_{true}$ is the true model parameters, $\boldsymbol{\epsilon}_d$ is the measurement errors which is usually assumed to follow a normal distribution $\mathcal{N}(0, \mathbf{C}_d)$ and $\mathbf{C}_d$ is the measurement errors covariance matrix. In this study, uncorrelated measurement errors are considered, therefore the matrix $\mathbf{C}_d$ is a diagonal matrix. In standard history matching (i.e. neglecting modeling errors), $\mathbf{C}_D$ in Eqs. 3, 4, 5 and 6 is set to the covariance of measurement errors. Therefore,

$$\mathbf{C}_D = \mathbf{C}_d. \tag{8}$$

However, as noted in the introduction section, several approximations are commonly introduced in the computational model to simplify the simulation process (e.g. black-oil model versus compositional flow), or to speed-up the simulations (coarsening of the simulation grid). During history matching if the model-error caused by these approximations is not accounted for, the obtained posterior distribution could be biased. In the case of utilizing an approximate/low-fidelity model, the observation data is related to the true model parameters $\mathbf{m}_{true}$ as (Giudice et al., 2013):

$$\mathbf{d}_{obs} = \tilde{g}(\mathbf{m}_{true}) + \boldsymbol{\epsilon}_d + \boldsymbol{\epsilon}_m, \tag{9}$$

where $\boldsymbol{\epsilon}_m$ is the model-error and $\tilde{g}(.)$ is a nonlinear function representing the imperfect (approximate/low-fidelity) simulation model. By subtracting Eq. 7 from 9, we obtain:

$$\boldsymbol{\epsilon}_m = g(\mathbf{m}_{true}) - \tilde{g}(\mathbf{m}_{true}). \tag{10}$$

In the following sub-section, we present a simple yet general parameterization of the model-error term $\boldsymbol{\epsilon}_m$.

2.2 Error-model formulation

In this study, EBD and I/O independent error-model approach is considered. As the model errors in our test cases were dominated by structured components, these errors are parameterized using smooth basis functions obtained by principle component analysis (PCA) method, which is an effective data-driven dimension reduction technique (Shlens, 2014; Kerschen et al., 2005). We rely on simulation output from pairs of models, accurate/high-fidelity versus approximate/low-fidelity, to obtain the basis functions and the prior statistics of the coefficients of the PCA basis functions. We acknowledge that this limits the applicability of the developed approach to the cases for which an accurate/high-fidelity model is available. However, we note that the accurate/high-fidelity model is only used to estimate the prior model-error statistics, and is not used during the calibration process. This is a notable difference between the presented framework and related studies by Josset et al. (2015) and Köpke et al. (2017).

In the current setting, the prior model-error realizations are estimated using:

$$\boldsymbol{\epsilon}_{mr} = g(\mathbf{m}_r) - \tilde{g}(\mathbf{m}_r), \tag{11}$$

where r is the index of the prior realizations, i.e. $r = 1$ to N_r, N_r is the total number of realizations used to estimate the model-error statistics, $\boldsymbol{\epsilon}_{mr}$ is an N_d dimensional vector of model-error for realization r. All prior model-error realizations are assembled into the matrix $\boldsymbol{\epsilon} \in \mathbb{R}^{N_d \times N_r}$. The mean of the model-error prior is,

$$\bar{\boldsymbol{\epsilon}}_m = \frac{1}{N_r} \sum_{r=1}^{N_r} (\boldsymbol{\epsilon}_{mr}). \tag{12}$$

The covariance of the model-error prior is (Oliver et al., 2008),

$$\mathbf{C}_e = \frac{1}{N_r - 1} (\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}_m \mathbf{I}_{N_r})(\boldsymbol{\epsilon} - \bar{\boldsymbol{\epsilon}}_m \mathbf{I}_{N_r})^\intercal, \tag{13}$$

where $\mathbf{I}_{N_r}$ is an N_r dimensional row vector with all ones as its components. In this study, PCA is used to parametrize the prior model-error realizations and the weights of the obtained PCA basis vectors are jointly inferred with the model parameters during the history matching process. The basis functions are obtained by singular value decomposition (SVD) of the error covariance matrix $\mathbf{C}_e$ (Oliver et al., 2008):

$$\mathbf{C}_e = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\intercal}, \tag{14}$$

where $\mathbf{U}$ and $\mathbf{V}$ are the orthonormal singular vectors (basis functions) and $\boldsymbol{\Sigma}$ is a diagonal matrix of the singular values. The error-model is formulated using the leading L singular vectors as following:

$$\hat{\boldsymbol{\epsilon}}_{mr} = \boldsymbol{\Phi}\boldsymbol{\beta}_r + \bar{\boldsymbol{\epsilon}}_m, \tag{15}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{N_d \times L}$ are the first L orthonormal singular vectors (basis functions) from $\mathbf{U}$ and $\boldsymbol{\beta}_r \in \mathbb{R}^{L \times 1}$ are the coefficients of error-model for realization r.

The objective of the calibration process is then to find the posterior distribution of reservoir model parameters and the coefficients $\boldsymbol{\beta}$ of the PCA-based error-model. Since Bayesian inverse modelling require prior statistics of model parameters, therefore the prior statistics of the coefficients $\boldsymbol{\beta}$ should be estimated. Least square form of Eq. 15 is used to compute prior realizations of the coefficient vector $\boldsymbol{\beta}$ as following:

$$\boldsymbol{\beta}_r = (\boldsymbol{\Phi}^{\intercal}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\intercal}(\boldsymbol{\epsilon}_{mr} - \bar{\boldsymbol{\epsilon}}_m). \tag{16}$$

Since $\boldsymbol{\Phi}$ is an orthonormal matrix i.e. $\boldsymbol{\Phi}^{\intercal} \approx \boldsymbol{\Phi}^{-1}$, therefore Eq. 16 can also be written as:

$$\boldsymbol{\beta}_r \approx \boldsymbol{\Phi}^{\intercal}(\boldsymbol{\epsilon}_{mr} - \bar{\boldsymbol{\epsilon}}_m). \tag{17}$$

The prior statistic, such as the mean and covariance of $\boldsymbol{\beta}$ realizations are computed using:

$$\boldsymbol{\mu}_{\beta} = \frac{1}{N_r}\sum_{r=1}^{N_r}(\boldsymbol{\beta}_r), \tag{18}$$

$$\mathbf{C}_\beta = \frac{1}{N_r - 1}\sum_{r=1}^{N_r}(\boldsymbol{\beta}_r - \boldsymbol{\mu}_\beta)(\boldsymbol{\beta}_r - \boldsymbol{\mu}_\beta)^\intercal, \tag{19}$$

where $\boldsymbol{\mu}_\beta \in \mathbb{R}^{L\times 1}$ is the mean of $\boldsymbol{\beta}$ realizations and $\mathbf{C}_\beta \in \mathbb{R}^{L\times L}$ is the covariance of $\boldsymbol{\beta}$ realizations. In this study, we only consider the diagonal terms of the matrix $\mathbf{C}_\beta$ to generate prior samples of error-model coefficients for the history matching purpose.

In order to avoid over-fitting the error-model, the number of coefficients of the PCA-based error-model L should be limited to a small number. Therefore, the residual of error-model cannot be neglected and need to be included in the inversion process. The residual of the least square fit is defined as,

$$\boldsymbol{\zeta}_{mr} = \boldsymbol{\epsilon}_{mr} - \hat{\boldsymbol{\epsilon}}_{mr}. \tag{20}$$

All residual realizations are assembled into matrix $\boldsymbol{\zeta} \in \mathbb{R}^{N_d\times N_r}$. The covariance of the residual from all error-model realizations is then estimated using:

$$\mathbf{C}_T = \frac{1}{N_r - 1}\boldsymbol{\zeta}\boldsymbol{\zeta}^\intercal, \tag{21}$$

where $\mathbf{C}_T$ is denoted the error-model noise covariance. For history matching of imperfect-models, the total error covariance matrix $\mathbf{C}_D$ in Eqs. 3, 4, 5 and 6 contains both the measurement and error-model noise components as following:

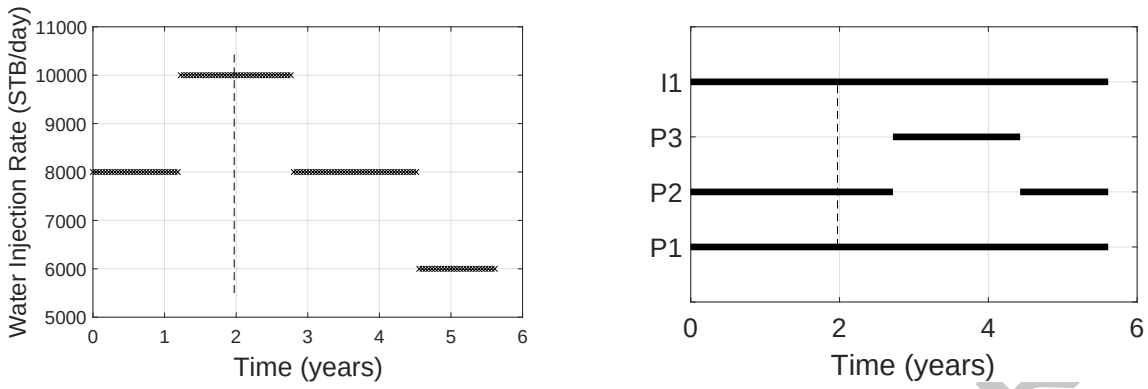
$$\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_T. \tag{22}$$

For simplicity, only the diagonal terms of the matrix $\mathbf{C}_T$ is considered in this study. Conceptually, $\mathbf{C}_D$ is the total uncalibrated uncertainty, which includes both the measurement noise and the model-error noise that is not captured by the truncated PCA-based error-model. In (Hansen et al., 2014), modeling errors were considered as uncalibrated uncertainties using Gaussian distribution (i.e. accounting for mean and covariance of errors). In their approach, the model error is accounted for by replacing $\mathbf{d}_{obs}$ with $\mathbf{d}_{obs} + \bar{\boldsymbol{\epsilon}}_m$ and replacing $\mathbf{C}_d$ with $\mathbf{C}_d + \mathbf{C}_e$. However, this approach would be inconsistent/inefficient for physical systems which exhibit highly complex statistics and

correlations of model errors that change significantly not only over the data space, but also as a function of the input model parameters. (Köpke et al., 2017).

3 Case Studies

In this section, we present the details of the case studies. The dimension of the subsurface reservoir is 7500 ft $\times$ 7500 ft $\times$ 20 ft in the x, y and z directions, respectively. Incompressible two-phase porous media flow of oil and water is considered. The initial reservoir pressure is 5000 psi and the reservoir has uniform porosity of 20%. The reservoir contains one injector well (I1) and three production wells (P1, P2, P3) and is simulated using a 2D grid. We utilize the Matlab Reservoir Simulation Tool-box (MRST) (Lie, 2016) for the forward model simulations. Corey's power law model is used to represent relative permeabilities. Parameter values for the Corey's model and fluid properties are listed in Appendix A. The gravitational and capillary pressure effects are neglected. The production wells are operated under constant bottom hole pressure constraint of 4500 psi and the injector well is operated under constant injection rate constraint with varying control values as shown in Fig. 1(a). The wells open/shut schedule is shown in Fig. 1(b). Figure 1 also shows end of historical period (i.e. 2 years). In the historical period, the flow rates at the production wells and the bottom hole pressure of the injector well are used as the historical data for the calibration process. We also note that one of the production wells (P3) is only used in prediction phase in order to assess predictions from calibrated models on wells drilled in future development plans.



(a) (b)

Figure 1: Injection well control rates and wells open/shut schedule. Dashed black lines show end of historical period. Part (a) shows the water injection rate of injector well. Part (b) shows wells open/shut schedule. In part (b) solid back lines indicate the time periods when a well is open to flow.

3.1 Case 1: Coarse scale model

In the first case study, fine-grid/high-fidelity model uses a 2D grid with 75×75 cells. The distributed log-permeability fields are modeled as multivariate Gaussian with exponential covariance function:

$$\mathbf{c} = \sigma^2 \exp(-3(\frac{\mathbf{s}}{r_a})^{\gamma}), \tag{23}$$

where $\mathbf{s}$ is the lag distance and r_a, σ^2, γ are the correlation range, variance and exponent respectively (which are 35 cells, 1 and 1 respectively in this test case). The log-permeability field $\ln(\mathbf{K})$ is parameterized using PCA and only two leading basis functions are retained:

$$\ln(\mathbf{K}) = \overline{\ln(\mathbf{K})} + \sum_{b=1}^{N_w} \mathrm{w}_b \boldsymbol{\psi}_b, \tag{24}$$

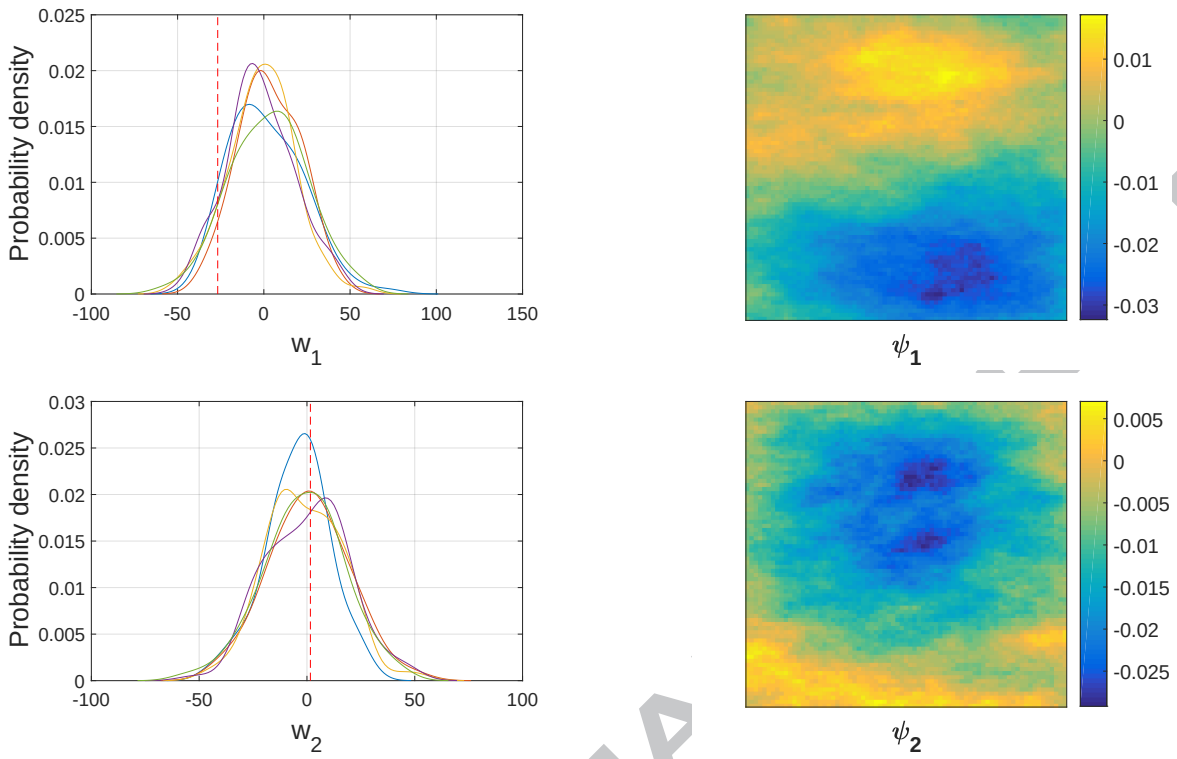
where $\overline{\ln(\mathbf{K})}$ is the mean log-permeability (equal to 4 in this test case), b is index of the basis weight w and basis function $\boldsymbol{\psi}$ and $N_w = 2$. Figure 2(b) shows the leading two principal basis functions obtained by singular value decomposition of covariance of log-permeability fields (Eq. 23). Figure 2(a) shows the prior distribution of weights obtained by projecting the log-permeability fields into the PCA-basis functions.

Figure 3(a) shows the reference fine scale log-permeability field. The fine scale reference log-

permeability field is generated by the leading two-PCA basis functions and reference basis weights are shown as the red vertical lines in Fig. 2. The coarse-grid/low-fidelity reservoir models contain only 5×5 grid blocks. The coarsened version of the reference fine model is shown in Fig. 3(b), in which harmonic averaging is used to up-scale the log-permeability field. The observed data are generated by the fine scale model using reference log-permeability field (Fig. 3(a)) with the addition of measurement noise of 2% of the reference solution. We note that except permeabilities, the rest of the static and dynamic properties (i.e. porosity, relative permeability, viscosity, and density, well controls and schedule) of the coarse scale model are the same as the fine scale model.

Two different procedures of history matching the coarse scale model were considered. In the case of neglecting model-discrepancy, PCA basis weights $\mathbf{w}$ of the log-permeabilities are calibrated (i.e. $\mathbf{m}_j = \mathbf{w}$ in Eq. 6). In the case of joint inversion with error-model, the estimated parameters is the combined vector of the log-permeability PCA weights and the error-model coefficients (i.e. $\mathbf{m}_j = [\mathbf{w}; \boldsymbol{\beta}]$ in Eq. 6).

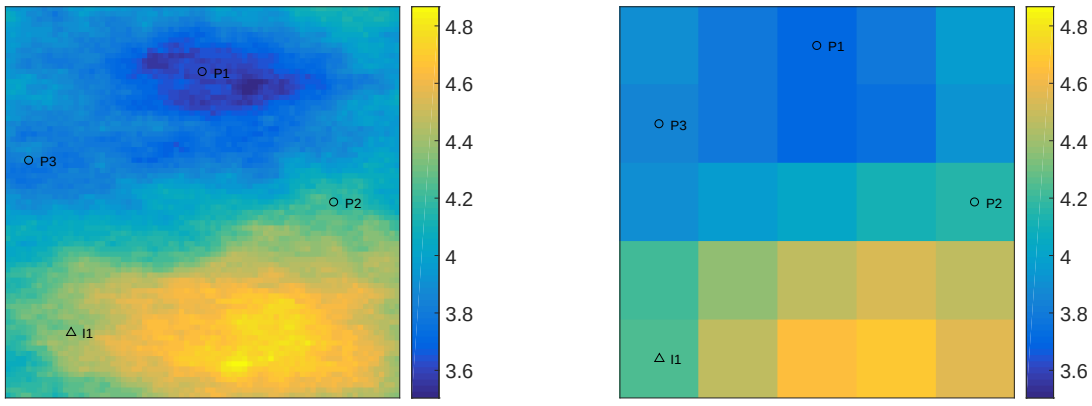
Prior statistics of the model-discrepancy were estimated using Eq. 11. One hundred fine scale permeability realizations were generated through Eq. 24 by sampling the prior distribution of the PCA-basis weights and a corresponding number of coarse scale permeability realizations were obtained using harmonic up-scaling. Forward runs were then performed for both the coarse and fine scale models to obtain the error realizations using Eq. 11. A smaller number of realizations could be used to evaluate the model-discrepancy statistics. In that case, special care should be taken to select a representative set of prior realizations to cover the respective statistic. Figure 4 shows the prior statistics of the model-discrepancy in the simulated well production data (bottom hole pressure of the injector well and flow rates of the producers).



(a) Prior distribution of basis weights

(b) Two principal basis functions

Figure 2: Prior distribution of basis weights from five ensembles and two principle basis functions for log-permeability. Red dashed lines show reference solution and five prior ensembles distribution are shown by five different colors in part (a).



(a)

(b)

Figure 3: The fine scale (75×75) reference log-permeability (a) and the corresponding up-scaled log-permeability (5×5) using harmonic average (b).

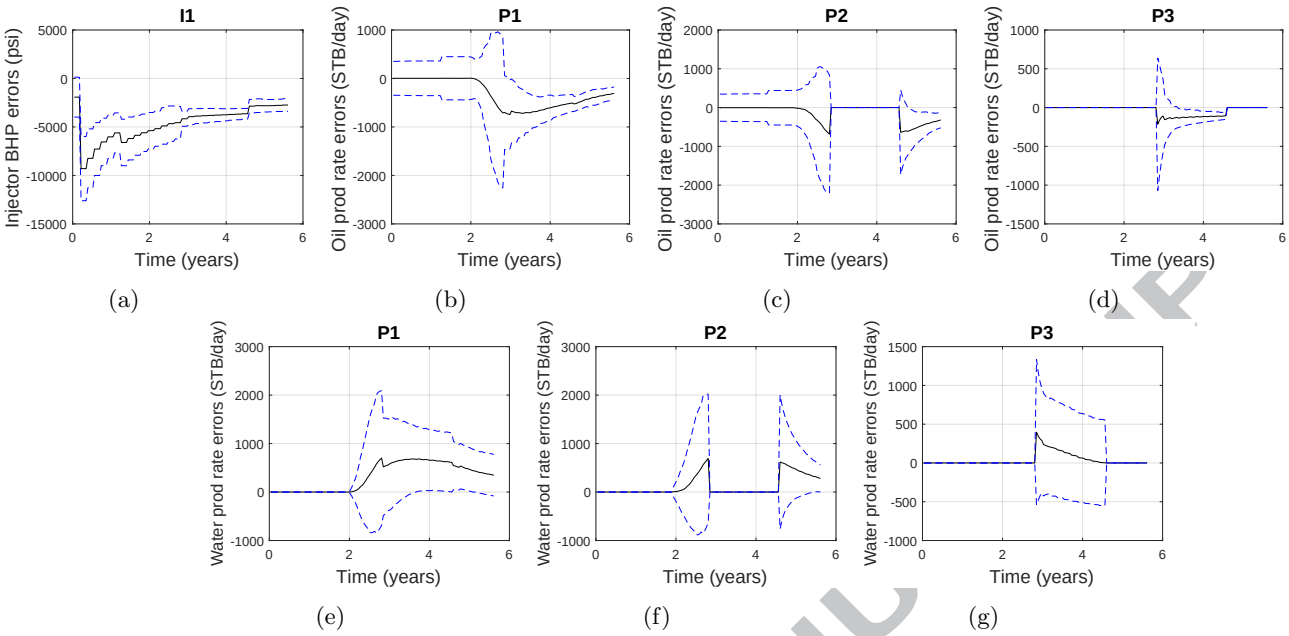


Figure 4: Prior model-error statistics of all wells for Case 1. Black lines show mean model errors, dashed blue lines show the 95% confidence interval (mean plus and minus two standard deviations) of model errors.

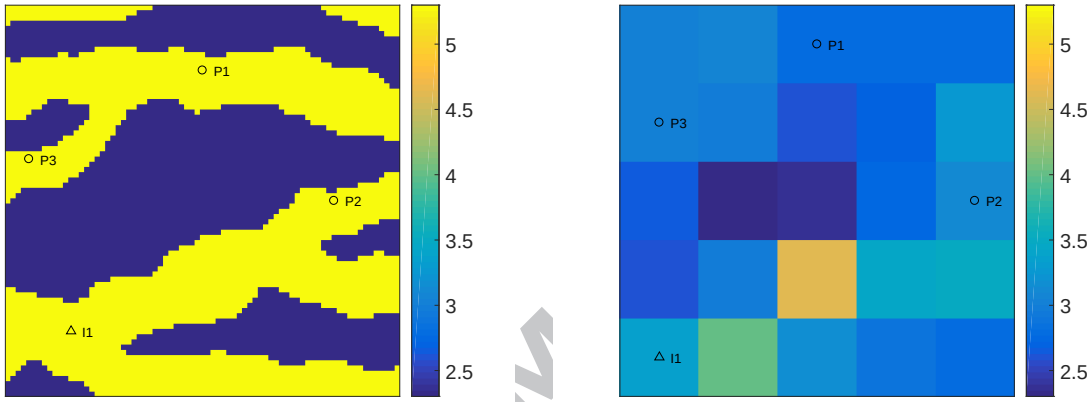
3.2 Case 2: Up-scaled imperfect geology model

Geologists commonly try to build geologically realistic prior models. However, maintaining the geological realism during the history-matching process is quite challenging (Sun and Durlofsky, 2017). For example, multipoint statistics (MPS) is widely used to represent channelized geological patterns. Geologically consistent history matching using MPS prior is still a subject of active research (Chen et al., 2016). Sometimes the predictability of the history matched MPS models may not be satisfactory, often due to limitation of the available history matching methods in handling this type of non-Gaussian models (Chen et al., 2016). In this study we do not aim for obtaining calibrated models that are consistent with the channelized geological feature, instead we focus on improving predictability of the calibrated coarse models by including the error-model.

For this case, the permeability fields are based on a similar test case presented in (Chen et al., 2016). Figure 5(a) shows the reference fine scale log-permeability with channelized features and Fig. 5(b) shows the corresponding up-scaled log-permeability field in which the channelized features have been lost due to harmonic averaging. The reference and prior fine scale channelized log-

permeability fields are generated using a two-facies training image with the direct sampling version of MPS (Mariethoz and Caers, 2014). The observed data are generated by the fine scale model using reference log-permeability field (Fig. 5(a)) with the addition of measurement noise of 2%.

Similar to the first test case, one hundred realizations of the model-discrepancy were obtained using Eq. 11 by running the fine scale simulation using the MPS permeability images of size 75×75 grid blocks and the corresponding up-scaled permeability field with 5×5 grid blocks. Figure 6 shows the prior statistics of the model-discrepancy in the simulated well production data.



(a) (b)

Figure 5: The fine scale (75×75) reference log-permeability with channelized features (a) and the corresponding up-scaled log-permeability (5×5) using harmonic average (b).

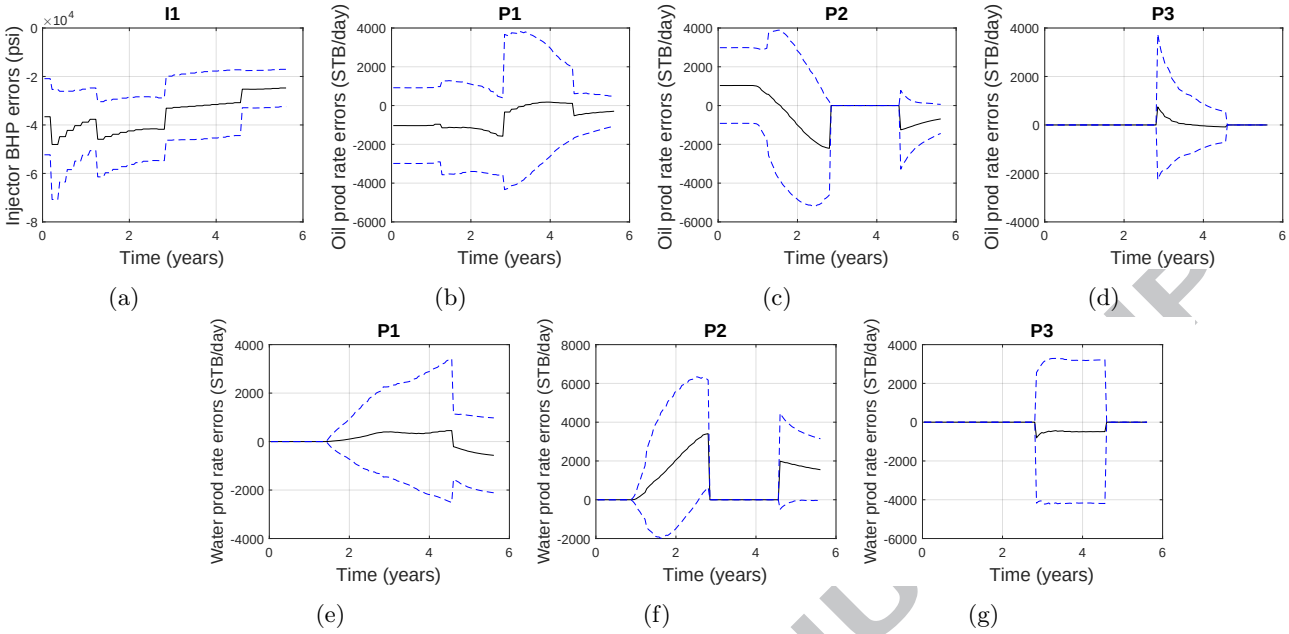


Figure 6: Prior model-error statistics of all wells for Case 2. Black lines show mean model errors, dashed blue lines show the 95% confidence interval (mean plus and minus two standard deviations) of model errors.

4 Results and Discussion

In this section, we present history matching results for test Case 1 and 2 with and without accounting for the model-discrepancy. All the ensemble-based history matching results are presented for multiple runs (five independent ensembles) in order to investigate the consistency and reliability of the parameter estimation process. Each ensemble run consists of 100 ensemble members and the measurement errors are assumed to be 2% of the observation data. We utilized the ES-MDA algorithm with eight iterations ($\alpha = 8$) for calibration. In the case of joint inversion with error-model, two PCA components were retained per each output time series to parametrize the model-discrepancy. Since we have seven output time series, the total number of error-model parameters is 14.

The calibrated models are evaluated using three different forecasting metrics to assess the quality of the estimated parameters and the capacity of the calibrated models in making future predictions. The utilized forecasting metrics are: coverage probability (CP), mean continuous ranked probability score (CRPS) and mean square error (MSE). CP indicates the fraction of the actual data that lie

within the 95% confidence interval of the estimation. A value of 0.95 for CP indicates a consistent estimation of uncertainty and values below 0.95 indicate underestimation of uncertainty. Mean CRPS quantifies both accuracy and precision (Hersbach, 2000) and higher values of CRPS indicate a less accurate results. MSE is widely used as a metric for parameter estimation problems. However, MSE measures the quality of data-fitting and is not enough to provide a probabilistic assessment of the estimation and prediction from an ensemble of models. In this study, we observed that a combination of MSE, CP and CRPS provides a good assessment of the quality for the probabilistic forecast (Skauvold and Eidsvik, 2018). The mathematical formulations of the three forecasting metrics are listed in Appendix B.

4.1 Case 1 Results

In this test case, the log-permeability is calibrated in terms of the PCA-basis weights, i.e. $\mathrm{w}_1, \mathrm{w}_2$ as detailed in the problem description. Figure 7 shows the posterior distributions of the basis weights for the five runs, for both cases of neglecting and accounting for model error. The results presented in Fig. 7(a) show that the posterior distribution from the inversion (neglecting model-error) procedure are biased and the estimated basis weights do not capture the reference weights. In contrast, the posterior distributions obtained by the joint inversion procedure are less biased (i.e. nearly unbiased) and successfully cover the true model parameters as shown in Fig. 7(b).

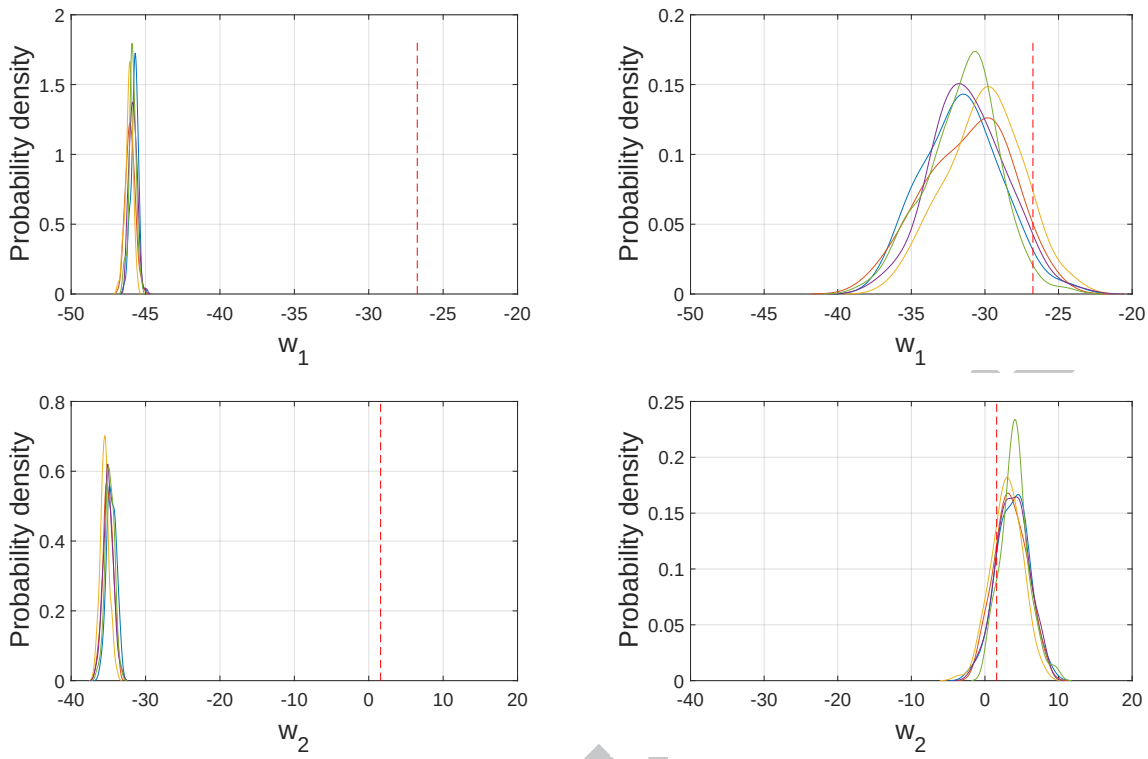
Figure 8 shows the mean and standard deviation of the posterior $\ln(\mathbf{K})$ for test case 1. In Fig. 8(a) the mean of posterior log-permeability field obtained from five different runs are shown for the inversion procedure. This posterior mean is clearly different from the coarse scale reference log-permeability field shown in Fig. 3(b) due to the bias in the inferred posterior distributions. Figure 8(b) shows the mean posterior log-permeability fields obtained by the joint inversion procedure. These fields are quite similar to the coarse scale reference log-permeability field. We also observe that the posterior standard deviations are quite low for the inversion (neglecting model-error) procedure, which could be a sign of over-fitting the data. In contrast for the joint inversion procedure, the standard deviations of the posterior fields are much larger due to accounting for the model error and including the error-model noise covariance matrix $\mathbf{C}_T$ in Eq. 22.

Figures 9 and 10 show the oil and water production rates of the different production wells and the bottom hole pressure of the injector well. The 50th percentile $p50$ and 95% confidence interval (the shaded region) are obtained by combining results from all five runs. In part (a) of these figures, the results for the inversion (neglecting model-discrepancy) are presented and the results of the joint-inversion with error-model are shown in part (b). For the inversion (neglecting model-error) procedure, the results are mixed where the data are matched for some cases and not matched for others. For example in Fig. 9(a), the data match is quite good for the wells P1 and P2. However in Fig. 10(a) the data match for well I1 is not as good. Moreover, the predictions for all wells are inaccurate for the inversion (neglecting model-error) procedure. Furthermore, the prediction envelop is really narrow, resulting in invisible confidence interval in the plots, which shows over-confidence in the inaccurate predictions. In comparison, better matches and predictions are obtained by the joint inversion procedure as shown in Figs. 9(b) and 10(b).

Figure 11 shows the forecasting metrics (CP for the estimated log-permeabilities and mean CRPS, MSE, CP for the well data in history matching and prediction periods) for individual ensemble (E1 to E5) and for results from all five ensembles assembled together (denoted as "All" in the figure). These metrics provide a good assessment on the consistency, reliability and accuracy of forecasting capacity of the calibrated models. Figures 11(a) and 11(b) show the coverage probability of reservoir model parameters ($\ln(\mathbf{K})$), well data for both the history matching period and the forecasting period. Both the inversion (neglecting model-error) and the joint inversion results are shown. In Fig. 11(a) CP of $\ln(\mathbf{K})$ is zero for each individual ensemble (E1 to E5) as well as for all five ensembles combined meaning that none of the ensemble captures the true log-permeability using the inversion (neglecting model-error) procedure. In Fig. 11(b) CP of $\ln(\mathbf{K})$ is one for E1 to E4 as well as for the combined ensemble that means that four out of five runs managed to enclose the reference log-permeabilities completely when using the joint inversion procedure. For the inversion (neglecting model-error) procedure, the CP lies between 0.06–0.11 for the historical data and lies between 0.01–0.013 for the validation data (prediction), as shown in Fig. 11(a). For the joint inversion procedure, these values of CP of are increased to be between 0.29–0.31 and 0.77–0.82, respectively as shown in Fig. 11(b). Although the value of CP equal to one (higher than the

correct value of 0.95) for the estimated permeability field when the joint inversion procedure is used clearly indicates that the uncertainty of the permeability field is overestimated, the overall results still show reasonable improvement from the joint inversion procedure compared to the standard inversion (neglecting model-error) procedure.

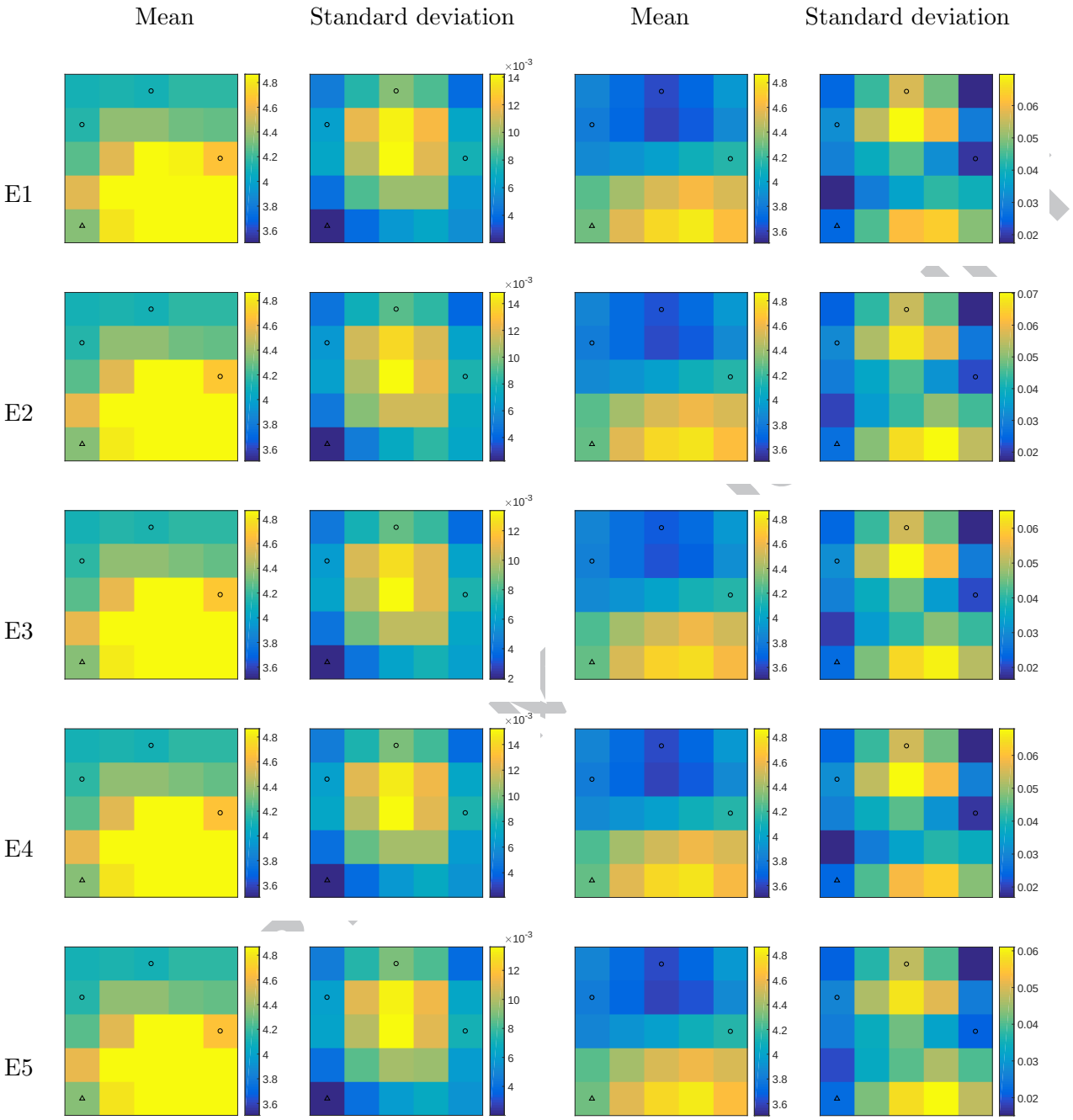
Figures 11(c) and 11(d) show the mean CRPS of history matching and prediction periods of the well data for the inversion (neglecting model-error) and the joint inversion procedures, respectively. Figure 11(c) shows that the mean CRPS lies between 117–118 and 466–469 for history matching and prediction periods, respectively. The results for the inversion (neglecting model-error) is unreliable and inaccurate due to the biased posterior distributions for all the different runs. Figure 11(d) shows that using joint inversion with error-model, the mean CRPS lies between 72–73 and 170–187 for history matching and prediction periods of the well data, respectively. A significant improvement in terms of reliability and accuracy is observed, by incorporating the error-model in the inversion process. Figures 11(e) and 11(f) show the MSE of the individual runs and the combined ensemble of all runs. With the joint inversion procedure, lower MSE values are obtained for both the history matching and the prediction periods (indicated by subscript "h" and "p" respectively in the plot).



(a) inversion (neglecting model-error)

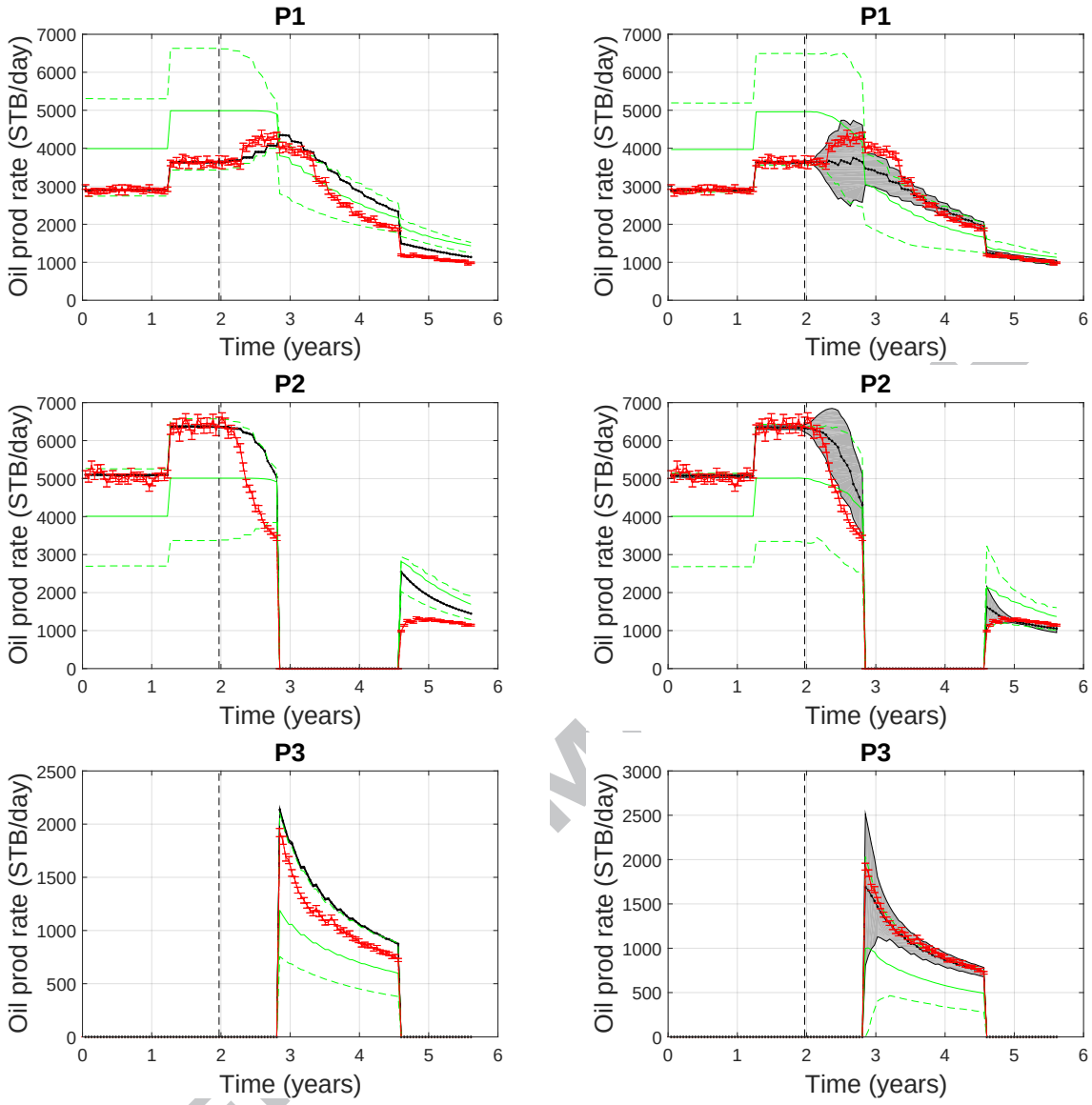
(b) joint inversion with error-model

Figure 7: Posterior distribution of two PCA basis weights (of $\ln(\mathbf{K})$) obtained after history matching for coarse scale model case. Dashed red lines show the reference solution and the posterior distribution of the five ensembles are shown by five different colors.



(a) inversion (neglecting model-error) (b) joint inversion with error-model

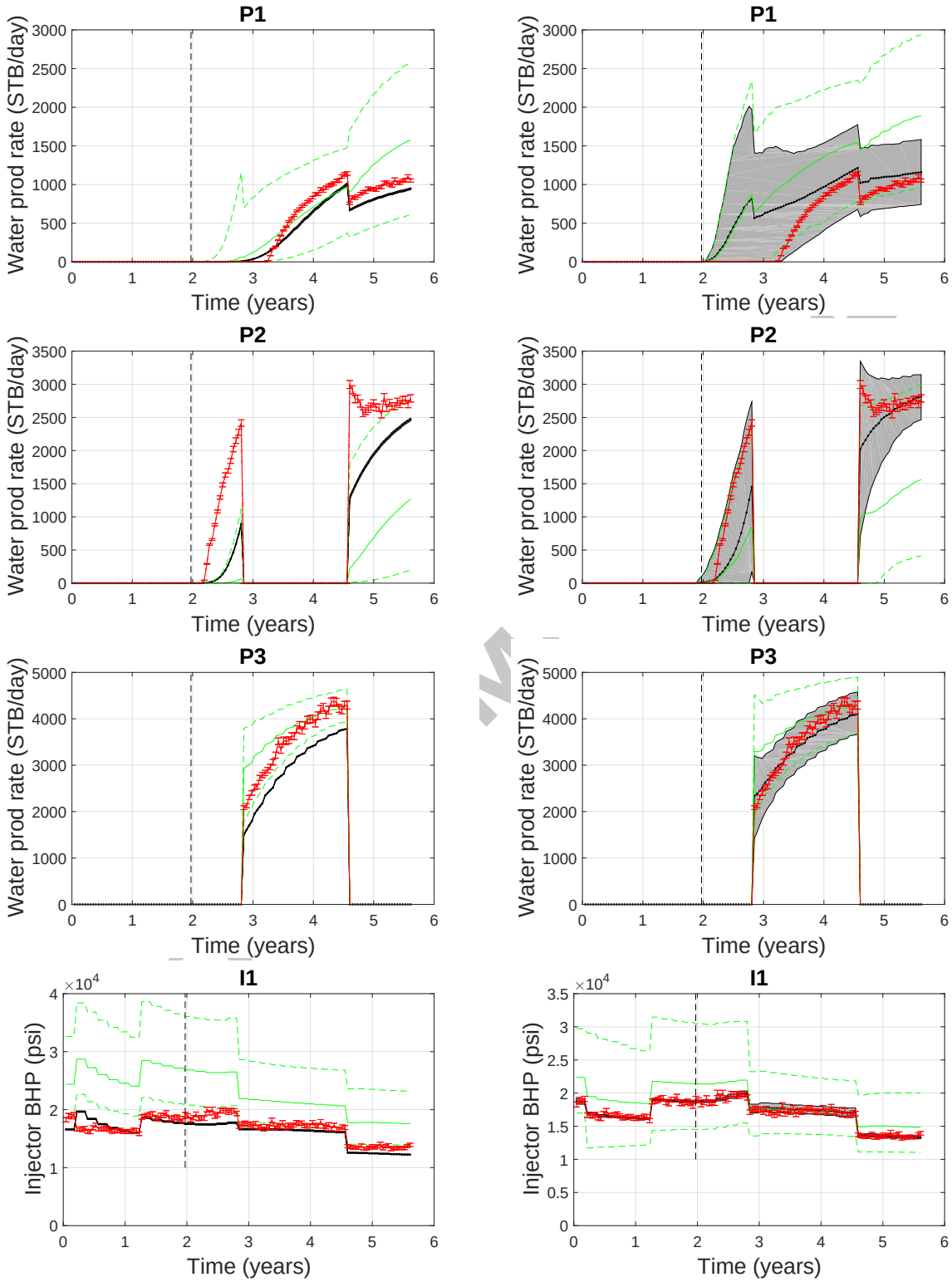
Figure 8: Mean and standard deviation of ln(**K**) posterior ensembles obtained after history matching of two PCA basis weights for coarse scale model case.



(a) inversion (neglecting model-error)

(b) joint inversion with error-model

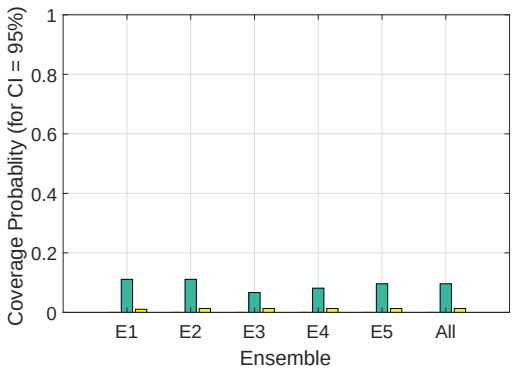
Figure 9: Prior and posterior of oil production data obtained from all ensembles for coarse scale model case. Red lines show observation data and bar on red lines shows measurement error. Dashed black lines show end of historical period. Solid green lines show 50th percentile $p50$ of prior distribution, dashed green lines show 95% confidence interval of prior distribution. Solid black lines show $p50$ of posterior distribution, gray shaded area shows 95% confidence interval of posterior distribution.



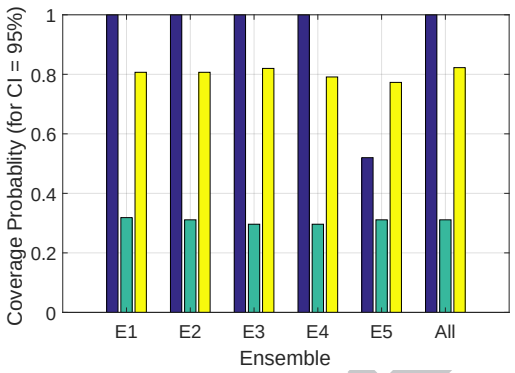
(a) inversion (neglecting model-error)

(b) joint inversion with error-model

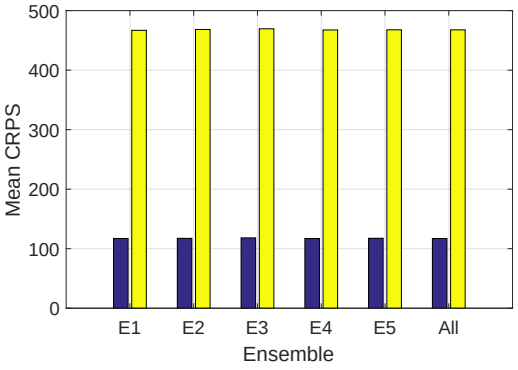
Figure 10: Prior and posterior of water production and injection pressure data for coarse model case. The explanation of colors and lines are the same as in Fig. 9.



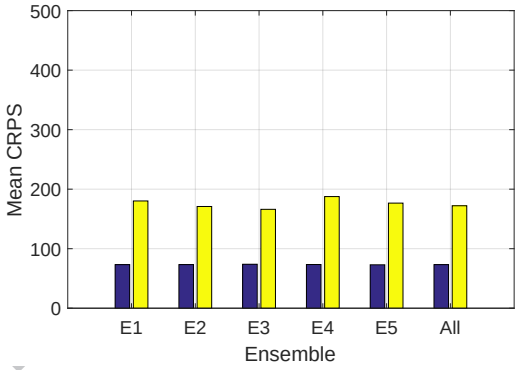
(a) inversion (neglecting model-error)



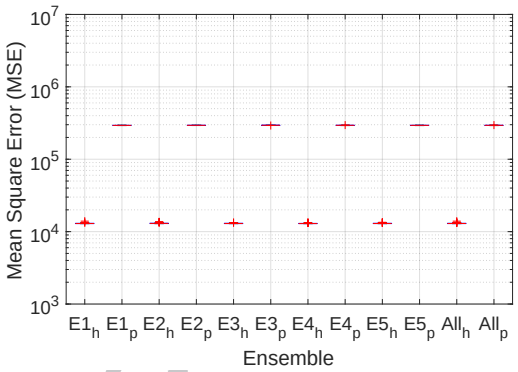
(b) joint inversion with error-model



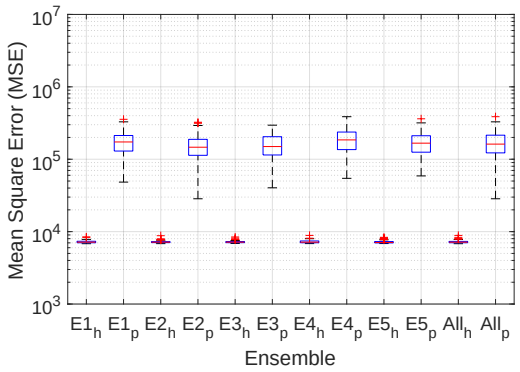
(c) inversion (neglecting model-error)



(d) joint inversion with error-model



(e) inversion (neglecting model-error)



(f) joint inversion with error-model

Figure 11: Forecasting metrics of coarse scale model case. In part (a) and (b) blue bars show the CP of true log-permeabilities, green bars show the CP of the historical data and yellow bars show the CP of prediction. In part (c) and (d) blue bars show the mean CRPS of the historical data and yellow bars show the mean CRPS of prediction. In part (e) and (f) box plots of MSE of the simulated well data from each ensemble are shown, subscript h and p are used for history and prediction respectively. On each box, the central red line indicates the median, and the bottom and top blue edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers represent extreme data points without outliers, and '+' symbol represents outliers (more than 1.5 times of interquartile range).

4.2 Case 2 Results

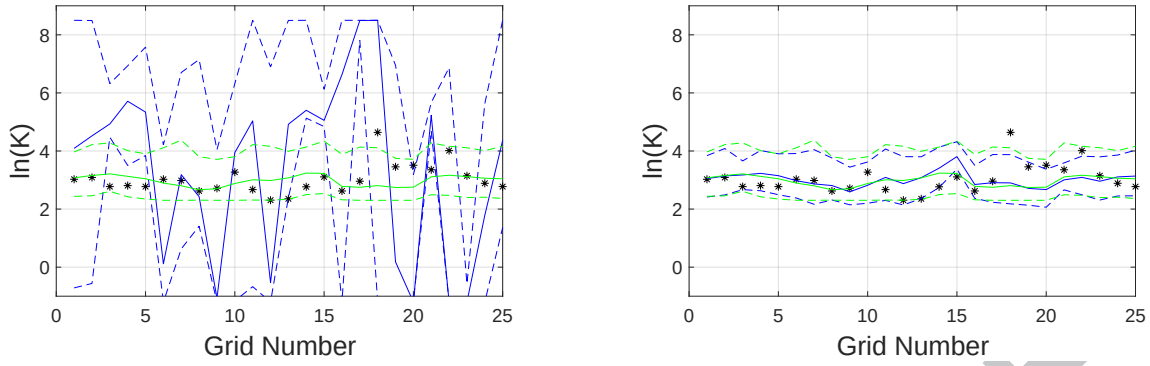
In this test case, log-permeability of every grid cell is calibrated using both inversion (neglecting model-error) and joint inversion procedures. Figure 12 shows prior and posterior distributions obtained by combining realizations from all five runs for both history matching procedures. Figure 12(a) shows that the posterior distribution of $\ln(\mathbf{K})$ is biased where the estimated log-permeabilities show extreme values and do not capture the reference log-permeabilities. Figure 12(b) shows that by accounting for model error, relatively small changes have been made to the physical parameters (log-permeability in this case), and the mean of the posterior distribution of $\ln(\mathbf{K})$ remains smooth after data assimilation.

Figure 13 shows the mean and standard deviation of $\ln(\mathbf{K})$ posterior ensembles as maps. In Fig. 13(a) the posterior mean log-permeability field obtained from five different runs are shown for the inversion (neglecting model-error) procedure. This posterior mean is clearly different from the reference coarse log-permeability field as shown in Fig. 5(b). Moreover, in Fig. 13(a) the mean log-permeability field of every ensemble is different from each other, which is an indication of convergence of every ensemble to different non-unique local peak (mode) of the biased posterior. The mean of log-permeability ensemble from the joint inversion procedure is shown in Fig. 13(b) where no extreme features are observed. Similar to the observation from Case 1, the standard deviation is higher for the joint inversion compared to the inversion (neglecting model-error).

Figures 14 and 15 show the oil and water production rates of the different production wells and the bottom hole pressure of the injector well. The 50th percentile $p50$ and 95% confidence interval (the shaded region) are obtained by combining results from all five ensembles. In part (a) of these figures, the results for the inversion (neglecting model-discrepancy) are presented and the results of the joint-inversion with error-model are shown in part (b). For the inversion (neglecting model-error) procedure, only historical data at some wells are matched. For example Figs. 14(a) and 15(a) show that the data match is good for wells I1 and P1, however the data match of well P2 is not good. Moreover the future prediction from the estimated model parameters using the inversion (neglecting model-error) procedure is quite poor. A good example is the prediction of water production rate of well P1: the models predicted early water breakthrough between year 2 to

3, while the actual water breakthrough is after year 5 because P1 is separated from the injector by a low permeability region (see Fig. 5(a)). In comparison, better matches and predictions are obtained by the joint inversion procedure as shown in Figs. 14(b) and 15(b) for individual ensembles as well as all ensembles combined together. The prediction from the combined multiple ensembles may seem good for some well data for the case with the inversion (neglecting model-error) procedure, for example BHP pressure of well I1 in Fig. 15(a), even though the prediction from each individual ensemble is not good. This is often due to the fact that different ensemble converges to different local peaks (modes) of the biased posterior and the combined prediction from these multiple biased final ensembles happen to enclose the validation data.

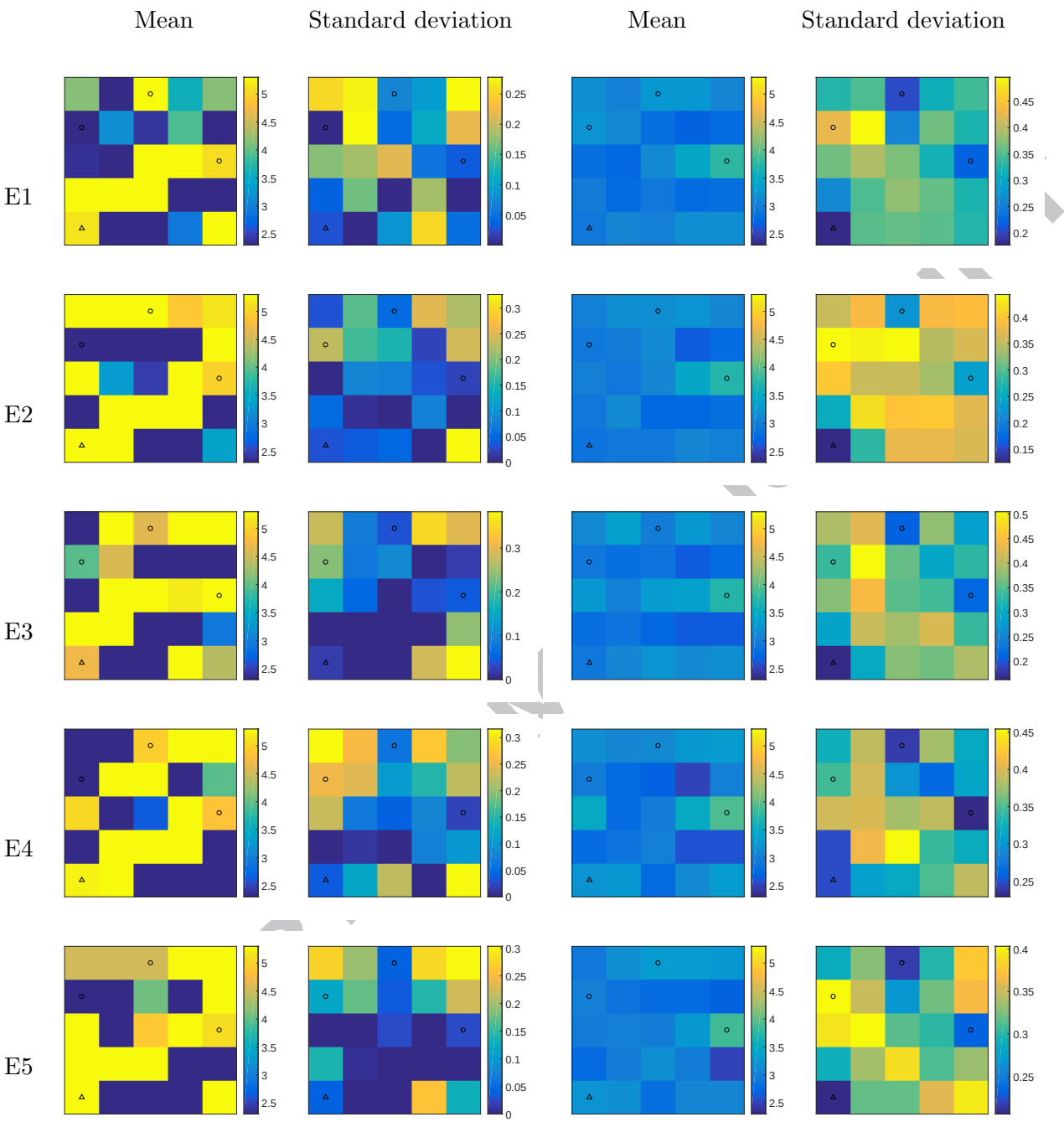
Figure 16 shows the forecasting metrics (CP for the estimated log-permeabilities and mean CRPS, MSE, CP for the well data in history matching and prediction periods) for the individual ensembles and for results from all five ensembles assembled together. Figures 16(a) and 16(b) show the coverage probability of reservoir model parameters ($\ln(\mathbf{K})$), well data for both the history matching period and the forecast period. In Fig. 16(a) CP of $\ln(\mathbf{K})$ is between 0–0.12 for ensembles (E1 to E5), however CP of $\ln(\mathbf{K})$ is 0.6 for the combined ensembles. This relatively high coverage from the combined ensemble is due to the overshooting of $\ln(\mathbf{K})$ values and the different final estimation from each individual ensemble as shown in Fig. 12(a) and Fig. 13(a) respectively. With the joint inversion procedure, the coverage probability is improved for all three quantities investigated (log-permeabilities, historical data, future prediction). Similarly, the mean CRPS and MSE measures also show significant improvement by accounting for model error using the joint inversion procedure (2nd and 3rd row of Fig. 16). In addition, based on all three forecasting measures, the results from multiple ensemble runs using the joint inversion procedure are very consistent, which indicates the statistical consistency of the proposed procedure.



(a) inversion (neglecting model-error)

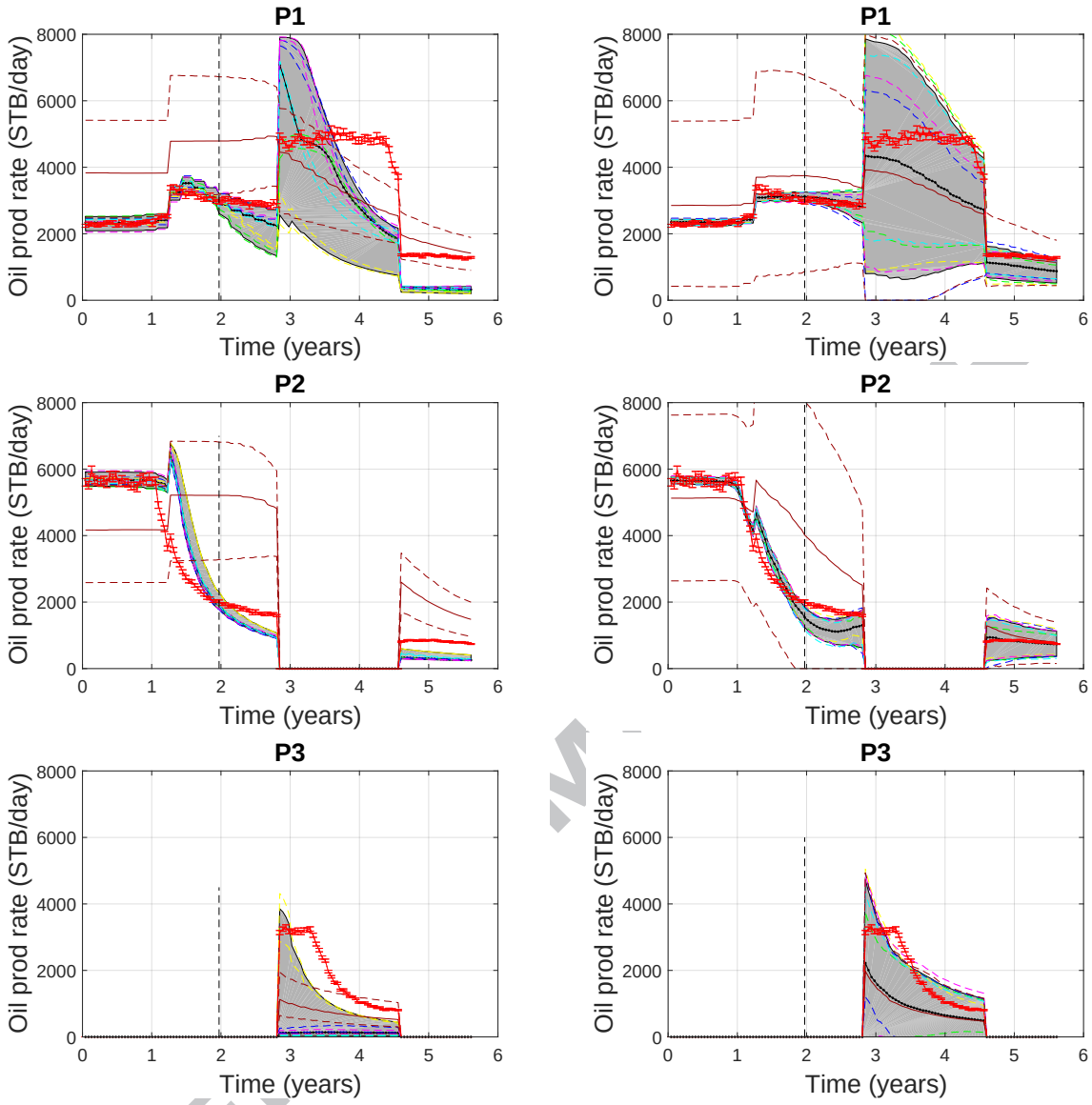
(b) joint inversion with error-model

Figure 12: Prior and posterior distribution of ln(**K**) obtained after history matching for up-scaled imperfect geology model case using five ensembles. In both part (a) and (b), green and blue lines show the prior and posterior distribution respectively. Solid green and blue line show the $p50$ prior and posterior respectively. Dashed green and blue lines show the 95% confidence interval of prior and posterior respectively. Black asterisks show the reference solution.



(a) inversion (neglecting model-error) (b) joint inversion with error-model

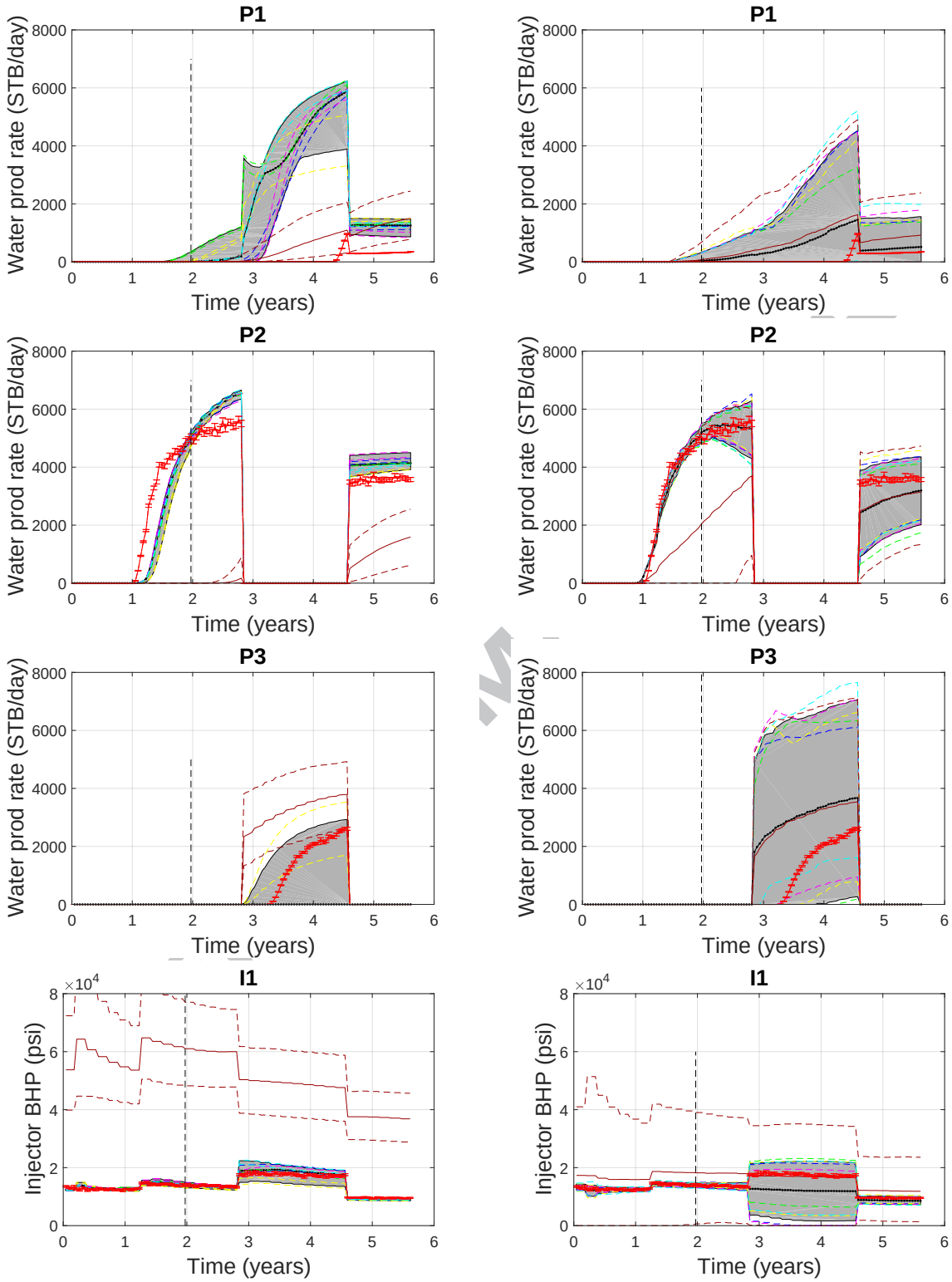
Figure 13: Mean and standard deviation of ln(**K**) posterior ensembles obtained after history matching of all grids log-permeabilities for up-scaled imperfect geology model case.



(a) inversion (neglecting model-error)

(b) joint inversion with error-model

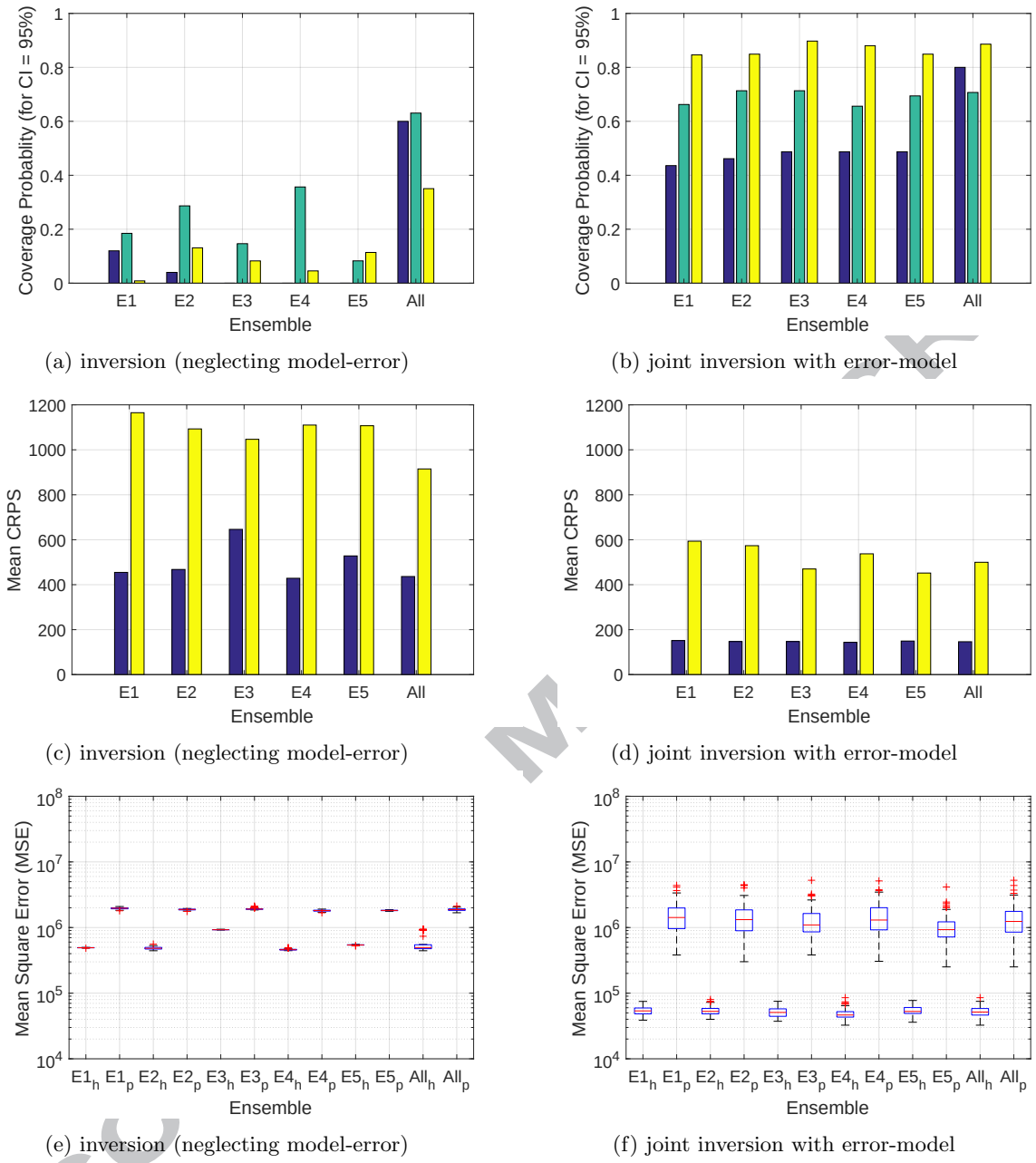
Figure 14: Prior and posterior of oil production data for up-scaled imperfect geological model case. Red lines show observation data and bar on red lines shows measurement error. Dashed black lines show end of historical period. Solid brown lines show 50th percentile $p50$ of the prior distribution, dashed brown lines show 95% confidence interval of prior distribution. Solid black lines show $p50$ posterior distribution obtained from all ensembles. Shaded gray area show 95% confidence interval of posterior distribution obtained from all ensembles. Dashed blue, green, yellow, magenta and cyan lines show 95% confidence interval of posterior distribution of individual ensembles.



(a) inversion (neglecting model-error)

(b) joint inversion with error-model

Figure 15: Prior and posterior of water production and injection pressure data for up-scaled imperfect geological model case. The explanation of colors and lines are the same as in Fig. 14.



(a) inversion (neglecting model-error)

(b) joint inversion with error-model

(c) inversion (neglecting model-error)

(d) joint inversion with error-model

(e) inversion (neglecting model-error)

(f) joint inversion with error-model

Figure 16: Forecasting metrics of up-scaled imperfect geology model case. In part (a) and (b) blue bars show the CP of true log-permeabilities, green bars show the CP of the historical data and yellow bars show the CP of prediction. In part (c) and (d) blue bars show the mean CRPS of the historical data and yellow bars show the mean CRPS of prediction. In part (e) and (f) box plots of MSE of the simulated well data from each ensemble are shown, subscript h and p are used for history and prediction respectively. On each box, the central red line indicates the median, and the bottom and top blue edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers represent extreme data points without outliers, and '+' symbol represents outliers (more than 1.5 times of interquartile range).

5 Conclusions

In this paper, a generic procedure for history matching of imperfect/low-fidelity reservoir models has been developed where, we formulate the history matching problem as a joint inversion of reservoir model parameters and an error model parameters. We used principal component analysis to parameterize the error model, where the PCA basis function and prior statistics of the PCA basis weights were obtained using pairs of accurate and inaccurate models. We note that the accurate model is only used for defining the prior model-error statistics and during history matching only the imperfect/low-fidelity model is used.

We evaluated the proposed history matching procedure on low-fidelity models with modeling errors due to aggressive grid coarsening/up-scaling of the permeability field obtained from two-point statistics and low-fidelity models where the main source of error is the grid coarsening/up-scaling of a channelized geology. Detailed comparison were performed against standard history matching (inversion while neglecting model error). The obtained results show that the estimated model parameters are biased using standard history matching procedure in the presence of large modeling errors. Consequently the calibrated low-fidelity model predictions are unreliable and generally inaccurate. Utilizing the developed joint inversion procedure results in significant improvements in terms of the quality of the estimated parameters, the matching capacity to historical data and prediction accuracy/reliability of the calibrated low-fidelity models. This is attributed to a reduction (and in some cases elimination) of the bias in the estimated posterior distribution of the model parameters when we included a flexible error-model terms in the inversion process. The numerical test cases were assessed using three forecasting metrics and it was observed that the consistency of ensemble-based history matching technique was also improved by including the error-model terms in the inversion procedure. We argue that this observed consistency might be due to the elimination of multiple biased peaks (modes) in the posterior distribution of the model parameters once the error modeling terms are included in the formulations.

The proposed framework is generally flexible and could be applied to large scale models as the error-model formulation is I/O independent and the prior error-model parameters could be estimated before the history matching step. However, for general error-modeling an accurate model

may be missing or the sources of the modeling errors could be unknown. It is also possible that the fine/high-fidelity model (which is assumed to be perfect) is also biased. In these cases, the proposed methodology can only improve the parameter estimation and the prediction up to the fine/high-fidelity model accuracy. Addressing the effects of unknown modeling errors without relying on an accurate (high-fidelity)/approximate (low-fidelity) model pairs is the subject of our future work.

6 Acknowledgments

The first author thanks Total E&P, UK for the financial support. The authors acknowledge Total S.A. for authorizing the publication of this paper.

Appendix A: Reservoir properties

Corey model in form of power law is used to generate relative permeability data for the reservoir model. Mathematically Corey model in form of power law is written as follows.

$$k_{rw} = (\hat{S_w})^{n_w} k_w^0. \tag{A.1}$$

$$k_{ro} = (1 - \hat{S_w})^{n_o} k_o^0. \tag{A.2}$$

$$\hat{S_w} = \frac{S_w - S_{wc}}{1 - S_{or} - S_{wc}}. \tag{A.3}$$

The notations of above equations are described in MRST manual Lie (2016). The fluid data and corey relative permeability model parameters used in the reservoir model are shown in the Table A.1.

Table A.1: Reservoir fluid data and Corey relative permeability model parameters

Fluid properties		Corey relative permeability model parameters			
water viscosity	0.5 cp	S_{or}	0.2	k_o^0	1
oil viscosity	1 cp	S_{wc}	0.2	k_w^0	1
water density	1000 kg/m^3	n_w	2		
oil density	700 kg/m^3	n_o	2		

Appendix B: Forecasting metrics

B.1: Mean Square Error (MSE)

The Mean Square Error (MSE) is obtained using,

$$MSE = \frac{1}{N_d}\sum_{n=1}^{N_d}(d_n - d_{obs,n})^2, \tag{B.1}$$

where n is index of observation or model prediction at corresponding time.

B.2: Coverage Probability (CP)

$$CP = \frac{N_{CI}}{N_t}. \tag{B.2}$$

N_{CI} = Number of samples, parameters or observations in Confidence Interval

N_t = Total number of samples, parameters or observations

B.3: Continuous Ranked Probability Score (CRPS)

The details of CRPS for ensemble prediction system were described by Hans Herbach (2000) Hersbach (2000). In this section summary of CRPS is explained.

Mathematically CRPS can be defined as,

$$CRPS = \int_{-\infty}^{\infty}[p(x) - H(x - x_{obs})]^2 dx, \tag{B.3}$$

where $p(x) = \int_{-\infty}^{x}\rho(y)dy$ Cumulative distribution of quantity of interest, $H(x - x_{obs})$ = Heaviside

function (Step function) i.e.

$$H(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

For an ensemble system with N_e realizations, the CRPS can be written as follows,

$$CRPS = \sum_{i=0}^{N_e} c_i. \tag{B.4}$$

$$c_i = \alpha_i p_i^2 + \beta_i (1 - p_i)^2. \tag{B.5}$$

where $p_i = P(x) = i/N_e$, for $x_i < x < x_{i+1}$ (Cumulative distribution is a piece wise constant function).

$$\alpha_i = \begin{cases} 0 & \text{if } x_{obs} < x_i \\ x_{obs} - x_i & \text{if } x_i < x_{obs} < x_{i+1} \\ x_{i+1} - x_i & \text{if } x_{obs} > x_{i+1} \\ x_{obs} - x_{N_e} & \text{if } x_{obs} > x_{N_e} \\ 0 & \text{if } x_{obs} < x_1 \end{cases}$$

$$\beta_i = \begin{cases} x_{i+1} - x_i & \text{if } x_{obs} < x_i \\ x_{i+1} - x_{obs} & \text{if } x_i < x_{obs} < x_{i+1} \\ 0 & \text{if } x_{obs} > x_{i+1} \\ 0 & \text{if } x_{obs} > x_{N_e} \\ x_1 - x_{obs} & \text{if } x_{obs} < x_1 \end{cases}$$

References

Asher, M. J., Croke, B. F., Jakeman, A. J., and Peeters, L. J. (2015). A review of surrogate models and their application to groundwater modeling. *Water Resources Research*, 51(8):5957–5973.

518 Cardoso, M., Durlofsky, L., and Sarma, P. (2009). Development and application of reduced-order
519 modeling procedures for subsurface flow simulation. *International journal for numerical methods*
520 *in engineering*, 77(9):1322–1350.

521 Carter, J. N., Ballester, P. J., Tavassoli, Z., and King, P. R. (2006). Our calibrated model has poor
522 predictive value: An example from the petroleum industry. *Reliability Engineering & System*
523 *Safety*, 91(10):1373–1381.

524 Chen, Y., Lallier, F., and Moncorge, A. (2016). On uncertainty quantification of history matched
525 facies models. In *ECMOR XV-15th European Conference on the Mathematics of Oil Recovery*.

526 Dreano, D., Tandeo, P., Pulido, M., Ait-El-Fquih, B., Chonavel, T., and Hoteit, I. (2017). Es-
527 timating model-error covariances in nonlinear state-space models using kalman smoothing and
528 the expectation–maximization algorithm. *Quarterly Journal of the Royal Meteorological Society*,
529 143(705):1877–1885.

530 Durlofsky, L. J. (2003). Upscaling of geocellular models for reservoir flow simulation: a review
531 of recent progress. In *7th International Forum on Reservoir Simulation Bühl/Baden-Baden,*
532 *Germany*, pages 23–27. Citeseer.

533 Emerick, A. A. and Reynolds, A. C. (2013). Ensemble smoother with multiple data assimilation.
534 *Computers & Geosciences*, 55:3–15.

535 Ertekin, T., Abou-Kassem, J. H. K., Gregory, R., Turgay Ertekin, J. H., and Gregory, R. K. (2001).
536 *Basic applied reservoir simulation*. Number Sirsi i9781555630898.

537 Evin, G., Thyer, M., Kavetski, D., McInerney, D., and Kuczera, G. (2014). Comparison of joint
538 versus postprocessor approaches for hydrological uncertainty estimation accounting for error
539 autocorrelation and heteroscedasticity. *Water Resources Research*, 50(3):2350–2375.

540 Giudice, D. D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., and Rieckermann, J. (2013).
541 Improving uncertainty estimation in urban hydrological modeling by statistically describing bias.
542 *Hydrology and Earth System Sciences*, 17(10):4209–4225.

Giudice, D. D., Löwe, R., Madsen, H., Mikkelsen, P. S., and Rieckermann, J. (2015). Comparison of two stochastic techniques for reliable urban runoff prediction by modeling systematic errors. *Water Resources Research*, 51(7):5004–5022.

Hansen, T. M., Cordua, K. S., Jacobsen, B. H., and Mosegaard, K. (2014). Accounting for imperfect forward modeling in geophysical inverse problems—exemplified for crosshole tomography. *GEOPHYSICS*, 79(3):H1–H21.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570.

Josset, L., Demyanov, V., Elsheikh, A. H., and Lunati, I. (2015). Accelerating monte carlo markov chains with proxy and error models. *Computers and Geosciences*, 85:38 – 48.

Kerschen, G., Golinval, J.-c., VAKAKIS, A. F., and BERGMAN, L. A. (2005). The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: An overview. *Nonlinear Dynamics*, 41(1):147–169.

Köpke, C., Irving, J., and Elsheikh, A. H. (2017). Accounting for model error in bayesian solutions to hydrogeophysical inverse problems using a local basis approach. *Advances in Water Resources*.

Kristensen, N. R., Madsen, H., and Jørgensen, S. B. (2004). Parameter estimation in stochastic grey-box models. *Automatica*, 40(2):225–237.

Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., and Jacques, D. (2013). Efficient posterior exploration of a high-dimensional groundwater model from two-stage markov chain monte carlo simulation and polynomial chaos expansion. *Water Resources Research*, 49(5):2664–2682.

Lie, K.-A. (2016). *An Introduction to Reservoir Simulation Using MATLAB - User Guide for the Matlab Reservoir Simulation Toolbox (MRST)*. Sintef ICT, Department of Applied Mathematics.

Lodoen, O. P., Omre, H., Durlofsky, L. J., and Chen, Y. (2005). Assessment of uncertainty in reservoir production forecasts using upscaled flow models. In *Geostatistics Banff 2004*, pages 713–722. Springer.

Lødøen, O. P. and Tjelmeland, H. (2010). Bayesian calibration of hydrocarbon reservoir models using an approximate reservoir simulator in the prior specification. *Statistical Modelling*, 10(1):89–111.

Maier, H. R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L. S., Cunha, M. C., Dandy, G. C., Gibbs, M. S., Keedwell, E., Marchi, A., et al. (2014). Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. *Environmental Modelling & Software*, 62:271–299.

Mariethoz, G. and Caers, J. (2014). *Multiple-point geostatistics: stochastic modeling with training images*. John Wiley & Sons.

Moradkhani, H., DeChant, C. M., and Sorooshian, S. (2012). Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-markov chain monte carlo method. *Water Resources Research*, 48(12).

Oliver, D. S. and Alfonzo, M. (2018). Calibration of imperfect models to biased observations. *Computational Geosciences*, 22(1):145–161.

Oliver, D. S., Reynolds, A. C., and Liu, N. (2008). *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge University Press.

Omre, H., Lodoen, O. P., et al. (2004). Improved production forecasts and history matching using approximate fluid-flow simulators. *SPE Journal*, 9(03):339–351.

O'Sullivan, A. and Christie, M. (2005). Error models for reducing history match bias. *Computational Geoscience*, 9(2-3):125–153.

Rammay, M. H. and Abdulraheem, A. (2014). Automated history matching using combination of adaptive neuro fuzzy system (anfis) and differential evolution algorithm. In *SPE Large Scale Computing and Big Data Challenges in Reservoir Simulation Conference and Exhibition*. Society of Petroleum Engineers.

592 Refsgaard, J. C., Christensen, S., Sonnenborg, T. O., Seifert, D., Højberg, A. L., and Troldborg,
593 L. (2012). Review of strategies for handling geological uncertainty in groundwater flow and
594 transport modeling. *Advances in Water Resources*, 36:36–50.

595 Reichert, P. and Schuwirth, N. (2012). Linking statistical bias description to multiobjective model
596 calibration. *Water Resources Research*, 48(9). W09543.

597 Shlens, J. (2014). A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*.

598 Silva, P. C., Maschio, C., and Schiozer, D. J. (2007). Use of neuro-simulation techniques as proxies
599 to reservoir simulator: application in production history matching. *Journal of Petroleum Science*
600 *and Engineering*, 57(3):273–280.

601 Skauvold, J. and Eidsvik, J. (2018). Data assimilation for a geological process model using the
602 ensemble kalman filter. *Basin Research*, 30(4):730–745.

603 Stordal, A. S. and Elsheikh, A. H. (2015). Iterative ensemble smoothers in the annealed importance
604 sampling framework. *Advances in Water Resources*, 86:231 – 239.

605 Sun, L., Seidou, O., Nistor, I., and Liu, K. (2016). Review of the kalman-type hydrological data
606 assimilation. *Hydrological Sciences Journal*, 61(13):2348–2366.

607 Sun, W. and Durlofsky, L. J. (2017). A new data-space inversion procedure for efficient uncertainty
608 quantification in subsurface flow problems. *Mathematical Geosciences*, 49(6):679–715.

609 Vrugt, J. A. (2016). Markov chain monte carlo simulation using the dream software package:
610 Theory, concepts, and matlab implementation. *Environmental Modelling & Software*, 75:273–
611 316.

612 White, J. T., Doherty, J. E., and Hughes, J. D. (2014). Quantifying the predictive consequences of
613 model error with linear subspace analysis. *Water Resources Research*, 50(2):1152–1173.